УДК: 004.8:614.253:004.056 https://doi.org/10.15407/jai2025.04.010

М. Я. Головенко[1], **В. Б. Ларіонов**[2]
[1,2]Фізико-хімічний інститут імені О. В. Богатського Національної академії наук України, Україна
вул. Люстдодська дорога, 86, м. Одеса, 65080
[1]n.golovenko@gmail.com
[2]vitaliy.larionov@gmail.com
[1]https://orcid.org/0000-0003-1485-128X
[2]https://orcid.org/0000-0003-2678-4264

ІНФОРМАЦІЙНЕ ВІКНО ЯК МЕТОДОЛОГІЯ ОЦІНКИ СПІВВІДНОШЕННЯ БЕЗПЕКИ ТА ЕФЕКТИВНОСТІ МЕДИЧНИХ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ

M. Golovenko[1], **V. Larionov**[2]
[1,2]O.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Ukraine
86, Lustdorfska Road, Odesa, 65080
[1]n.golovenko@gmail.com
[2]vitaliy.larionov@gmail.com
[1]https://orcid.org/0000-0003-1485-128X
[2]https://orcid.org/0000-0003-2678-4264

INFORMATION WINDOW AS A METHODOLOGY FOR ASSESSING THE SAFETY AND EFFECTIVENESS BALANCE OF ARTIFICIAL INTELLIGENCE MEDICAL SYSTEMS

Анотація. У статті розглянуто методологічні підходи до оцінки безпеки та клінічної ефективності систем штучного інтелекту (ШІ) в охороні здоров'я. Проведено порівняльний аналіз концепції «користь–ризик» у фармакології та медичних ШІ-системах із метою адаптації концепції терапевтичного вікна до інформаційних технологій. Запропоновано модель інформаційного вікна як формалізованого інструменту для оцінювання співвідношення очікуваної користі та потенційних ризиків; під цим поняттям розуміється діапазон параметрів прийняття рішень, у межах якого система ШІ зберігає оптимальний баланс між безпекою та ефективністю. Враховано сучасні регуляторні вимоги, технічні показники ефективності, етичні аспекти та соціальні наслідки. Обґрунтовано необхідність етапного тестування ШІ-систем за аналогією до клінічних випробувань лікарських засобів. Окреслено перспективу розробки міжнародних нормативів для визначення допустимих індексів користі–ризику. Наголошено на важливості безперервного аудиту, оновлення моделей та встановлення механізмів відповідальності. Запропонований підхід сприяє формуванню уніфікованих стандартів для безпечного та ефективного впровадження ШІ в медичну практику.

Ключові слова: штучний інтелект, безпека, ефективність, терапевтичне вікно, інформаціне вікно, сучасні регуляторні вимоги.

Abstract. The article examines methodological approaches to assessing the safety and clinical efficacy of artificial intelligence (AI) systems in healthcare. A comparative analysis of the "benefit–risk" concept in pharmacology and medical AI systems is conducted to adapt the concept of the therapeutic window to information technologies. A model of the *information window* is proposed as a formalized tool for evaluating the balance between expected benefits and potential risks; this concept refers to the range of decision-making parameters within which an AI system maintains an optimal balance between safety and efficacy. Current regulatory requirements, technical performance indicators, ethical aspects, and social implications are taken into account. The necessity of phased testing of AI systems, analogous to clinical trials of pharmaceuticals, is substantiated. The prospects for developing international standards to define acceptable benefit–risk indices are outlined. The importance of continuous auditing, model updating, and establishing mechanisms of accountability is emphasized. The proposed approach contributes to the development of unified standards for the safe and effective implementation of AI in medical practice.

Keywords: artificial intelligence, safety, efficacy, therapeutic window, information window, modern regulatory requirements.

Вступ

Ми живемо в епоху четвертої промислової революції, коли автоматизація, штучний інтелект (ШІ) та обробка великих даних стають визначальними чинниками розвитку суспільства та економіки [1].

Охорона здоров'я є однією з тих галузей, де трансформаційний потенціал ШІ проявляється особливо яскраво. Сучасні алгоритми, від методів машинного навчання та обробки природної мови до роботизованої автоматизації процесів, відкривають нові можливості для прогнозного моделювання, раннього виявлення захворювань, оптимізації клінічних рішень і навіть підвищення точності хірургічних втручань. Поява великих мовних моделей, таких як BERT, GPT та їхні похідні, ще більше розширила застосування ШІ, зокрема у сфері клінічної документації та аналізу медичної літератури [2]. Така інтеграція ШІ в медицину обіцяє підвищення точності діагностики, зменшення кількості помилок, оптимізацію лікувальних стратегій і, як наслідок, покращення результатів для пацієнтів. Водночас регуляторний ландшафт перебуває на стадії активного становлення, прагнучи створити ефективні механізми контролю й управління впровадженням цих технологій у клінічну практику. Це особливо важливо з огляду на значні ризики, що супроводжують застосування ШІ, як-от алгоритмічну упередженість, загрозу конфіденційності, потенційні соціальні наслідки та неконтрольовані помилки, які у сфері охорони здоров'я можуть мати критичний характер.

Проблематика співвідношення «користь–ризик» є ключовою у будь-якій галузі науки й техніки. У фармакології ця концепція реалізується через ідею «терапевтичного вікна», що визначає діапазон між мінімальною ефективною та максимально безпечною дозою препарату [3,4]. Жоден лікарський засіб не є абсолютно безпечним і навіть такі поширені препарати як парацетамол чи аспірин, можуть спричиняти побічні ефекти. Тому лікарі завжди зважують, наскільки очікувана користь від лікування перевищує потенційні ризики. Історичні трагедії, шо стались з талідомідом, наочно показали небезпеку недооцінки ризиків та відсутності належного регуляторного контролю.

Медичні системи ШІ стикаються з подібними викликами. Як і ліки, вони мають потенціал приносити користь, але водночас несуть ризики, пов'язані з хибними рішеннями, упередженістю чи непрозорістю алгоритмів. Однак на відміну від фармакології, де концепція терапевтичного вікна стала фундаментом для оцінки безпеки та ефективності, у сфері медичного ШІ стандартизований підхід до системної оцінки співвідношення «користь-ризик» знаходиться на початковій стадії [5].

Таким чином, попри значний прогрес у розвитку методів оцінки ШІ в медицині, сучасний ландшафт залишається фрагментованим: одні підходи зосереджуються на продуктивності алгоритмів, інші на питаннях безпеки, етики чи регуляторної відповідності. Відсутність єдиного інтегрованого підходу ускладнює створення комплексної картини користі й ризиків медичних систем ШІ.

Постановка проблеми

Лікарські засоби і ШІ у медицині є інструментами втручання, які безпосередньо впливають на діагностику, лікування, прогнозування результатів пацієнтів. Обидва підходи не є «нейтральними»: вони можуть приносити значну користь, але й нести суттєві ризики. Саме тому порівняння користь–ризик є спільною основою для клінічного застосування. Для лікарських засобів джерелом користі є фармакологічна дія, що реалізується через взаємодію активної речовини з біохімічними мішенями (рецепторами, ферментами, іонними каналами тощо). У випадку систем ШІ користь формується завдяки алгоритмічній точності, здатності системи правильно інтерпретувати дані та підтримувати клінічні рішення (наприклад, прогнозування ризику, автоматична діагностика, персоналізація лікування). Таким чином, в обох випадках користь реалізується через вплив на медичне рішення, що обґрунтовує їхнє порівняння в межах єдиної моделі «користь–ризик».

Ризики у фармакотерапії, як правило, пов'язані з токсичністю, передозуванням або непередбаченими побічними реакці-

ями. Для ШІ-рішень характерними є наступні інформаційні ризики: хибні діагнози (false positive/false negative); алгоритмічна упередженість, bias), що призводить до нерівного доступу або помилкових рішень для різних груп пацієнтів та неможливість пояснити логіку прийнятого рішення. Обидва типи ризиків мають спільну властивість бо вихід за межі контрольованої дії може перевищити очікувану користь, що створює необхідність визначення «інформаційного» чи «терапевтичного» вікна безпеки.

У фармакології ефективність і безпека відстежуються через моніторинг рівня препарату в крові, клінічне спостереження побічних ефектів, фармаконагляд [6]. У ШІ цей процес реалізується у формі валідації моделі, постмаркетингового аудиту та тестування на упередженість [7]. Після впровадження системи в практику її продуктивність повинна перевірятися аналогічно моніторингу лікарського препарату у популяції пацієнтів. Таким чином, обидві системи потребують безперервного циклу оцінки безпеки, де ШІ виступає цифровим аналогом фармаконагляду.

Регуляторне поле для лікарських засобів визначається установами ЕМА (Європейське агентство з лікарських засобів), FDA (Управління з контролю якості харчових продуктів і медикаментів США), а також національними системами фармаконагляду (ДЕЦ МОЗ України). Для ШІ медичного призначення діють нормативні акти FDA SaMD Framework, EU AI Act (2024), а також Регламент MDR (Medical Device Regulation). Спільною основою для обох напрямів є вимога обґрунтування співвідношення користь–ризик, яке виступає ключовим критерієм допуску технології до ринку. Це підтверджує наявність спільної логіки регулювання між біомедичними й цифровими інноваціями. Однак на відміну від фармакології, де концепція терапевтичного вікна стала фундаментом для оцінки безпеки та ефективності ліків, у сфері медичного ШІ стандартизований підхід до системної оцінки співвідношення «користь–ризик» знаходиться на початковій стадії.

Наведене порівняння показує, що системи ШІ медичного призначення можна розглядати як функціональний аналог лікарських засобів у цифровій сфері. Їхня ефективність та безпека залежать від точності дозування інформації, контролю якості даних і постійного моніторингу наслідків застосування. Саме в цьому контексті концепція «інформаційного вікна» (ІВ) набуває ключового значення, що дозволяє кількісно оцінити баланс між користю та ризиком у цифровій медицині, подібно до «терапевтичного вікна» у фармакології. Розробка та впровадження концепції ІВ є не лише академічним завданням, а й необхідністю для безпечної та ефективної медицини майбутнього.

Така регульована інтеграція ШІ в медицину обіцяє підвищення точності діагностики, зменшення кількості помилок, оптимізацію лікувальних стратегій і, як наслідок, покращення результатів для пацієнтів.

Аналіз останніх досліджень та публікацій

Аналіз публікацій вітчизняних та зарубіжних учених показав, що на сьогоднішній день активно обговорюються і розвиваються ключові аспекти розвитку ШІ та його впливів на різні галузі суспільства, зокрема медицини [8-12]. Моделі ШІ дозволяють ефективно вирішувати завдання класифікації, прогнозування, оптимізації, управління економічними процесами, розробки соціальних програм та інші, що має важливе значення для багатьох областей, включаючи медицину, фінанси, промисловість тощо.

Інтеграція систем ШІ у сферу охорони здоров’я супроводжується низкою методологічних, організаційних та регуляторних викликів. Важливим завданням є формування довіри до систем ШІ серед медичних працівників і пацієнтів. Це передбачає не лише їх технічну валідність, але й етичну прийнятність, а також готовність до інтеграції у клінічні робочі процеси. Використання ШІ не

повинно обмежуватися роллю допоміжного інструмента, а оптимальним є його впровадження як невід'ємного елементу оновленої моделі надання медичної допомоги, що підвищує ефективність та якість лікування.

Ключовим чинником залишається забезпечення сталого фінансування інновацій, особливо у державних закладах охорони здоров'я, де впровадження ШІ може значно зменшити ресурсні витрати. Водночас необхідним є формування механізмів юридичної відповідальності, які дозволяють пацієнтам отримувати компенсацію у випадках шкоди, завданої дефектними алгоритмічними системами.

Законодавче регулювання відіграє визначальну роль у формуванні безпечного середовища для використання ШІ в медицині. У цьому контексті особливого значення набув Європейський акт про ШІ [5], що набрав чинності 1 серпня 2024 року. Документ встановлює такі особливі вимоги як мінімізація ризиків, використання якісних навчальних даних, забезпечення прозорості та зрозумілої інформації для користувачів, а також гарантування належного людського нагляду за процесами прийняття рішень.

Таким чином, інтеграція ШІ в охорону здоров'я вимагає комплексного підходу, який поєднує технічні, клінічні, етичні та правові аспекти. Тільки за умови збалансованого розвитку цих складових можливо досягти його повного потенціалу у покращенні результатів лікування та безпеки пацієнтів.

Мета та завдання дослідження

Метою нашого дослідження є створення нової концепції ІВ, аналога терапевтичного вікна у фармакології, яка може слугувати основою для систематичної оцінки ефективності й безпеки медичних ШІ. Цей підхід дає змогу не лише уніфікувати критерії безпеки й ефективності, а й забезпечити більш прозоре прийняття клінічних та регуляторних рішень у сучасній медичній практиці. Задля цього обґрунтувано та проведено порівняльний аналіз інтегральних показників співвідношення користі та ризику у фармакології та ШІ медичного призначення, з акцентом на можливості адаптації концепції терапевтичного вікна до розробки аналогічного індексу для інформаційних систем. З цією метою використано наукову літературу з клінічної фармації, зокрема монографії та керівні настанови щодо визначення терапевтичного індексу та терапевтичного вікна лікарських засобів. Враховано основні позиції документів міжнародних та національних регуляторних органів, що регламентують критерії безпеки та ефективності фармакотерапії та нормативні ініціативи у сфері ШІ (AI Act, рекомендації OECD, принципи етики ШІ від IEEE).

Застосовано порівняльно-аналітичний метод для аналізу аналогій між терапевтичним вікном у фармакології та ІВ у ШІ. Систематизацію та класифікацію використано для визначення основних компонентів користі та ризику в обох сферах, а формалізацію для побудови моделі розрахунку ІВ здійснювали за допомогою спеціальних параметрів оцінки ефективності алгоритмів ШІ.

У такий спосіб, дослідження носить концептуально-теоретичний характер і сприяє формуванню узагальненої моделі оцінювання безпеки та ефективності ШІ.

Основний матеріал дослідження

Діапазон ефективності та безпеки лікарських засобів та перенесення концепту дозування на сферу інформаційних систем

Концепція терапевтичного вікна є ключовою у галузі фармакології та медичного лікування, яка стосується діапазону дозування препарату для створення бажаної дії, не спричиняючи неприйнятних побічних ефектів [4, 5]. По суті, терапевтичне вікно відображає баланс між ефективністю та безпекою, де нижній поріг цього вікна становить мінімальну ефективну дозу (MED), яка є найменшою кількістю препарату, необхідною для досягнення клінічно значущої відповіді. Вище цього рівня препарат надає бажані терапевтичні ефекти. Верхній поріг, з іншого боку, є максимальною безпечною

дозою (MSD), після якої препарат може викликати побічну дію, що є загрозою для здоров'я (рис. 1). Ширина терапевтичного вікна значно варіюється між різними препаратами. Препарати з вузьким терапевтичним вікном, між MED та MSD вимагають ретельного моніторингу та точного дозування для підтримки концентрації препарату в сироватці крові в межах безпечного та ефективного діапазону [6]. Такі препарати часто потребують регулярного моніторингу за допомогою аналізів крові, оскільки незначні відхилення від призначеної дози можуть призвести до серйозної токсичності або неефективності. Визначення терапевтичного діапазону ґрунтується не лише на фізіологічних реакціях, але й включає фармакокінетичні та фармакодинамічні властивості препарату, які описують, як організм впливає на препарат і навпаки [5]. Такі фактори, як вік, маса тіла, генетична схильність та наявність інших захворювань, можуть впливати на ці процеси та тим самим впливати на терапевтичне вікно.

Запропонована нами концепція ІВ може також бути ключовою у сфері оцінки та регуляції медичного ШІ. Вона стосується діапазону продуктивності та ризику застосування алгоритму, у межах якого система ШІ забезпечує клінічно значущий ефект без створення неприйнятних небезпек для пацієнтів чи медичної системи (рис. 1). Розуміння та визначення цього вікна є критично важливим для забезпечення балансу між користю та ризиком у використанні алгоритмів.

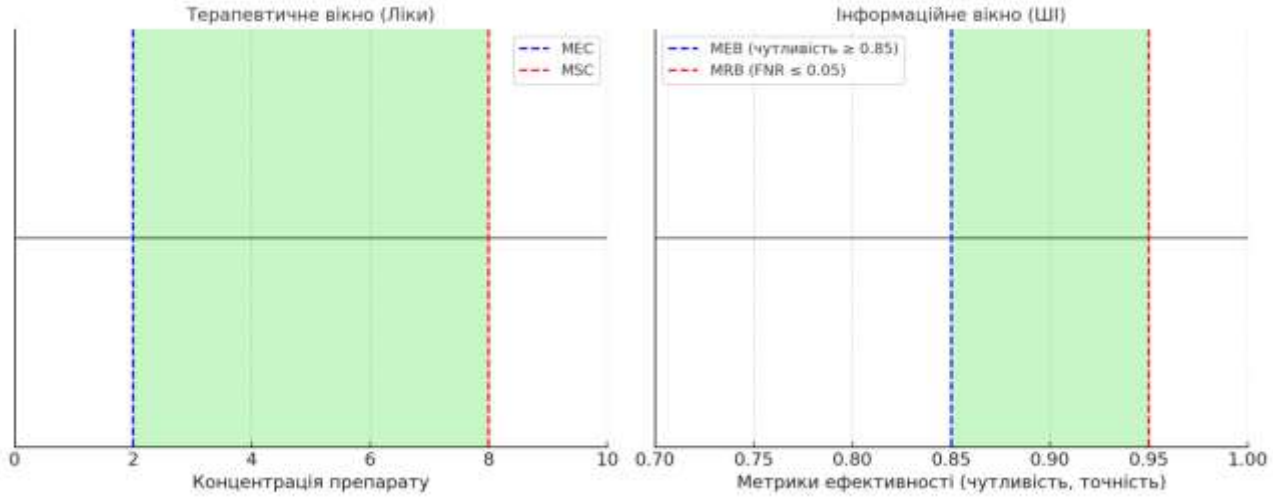

Рис. 1. Аналогія між фармакологічною концепцією терапевтичного вікна та новою концепцією ІВ для медичних ШІ

Ліки (зліва):

Синя лінія - мінімальна ефективна концентрація препарату. Нижче цього рівня ліки не дають терапевтичного ефекту.

Червона лінія - максимальна безпечна концентрація. Вище цього рівня починається токсичність і ризик для пацієнта.

Зелена зона є терапевтичним вікном, тобто діапазоном, де препарат ефективний і водночас безпечний.

ШІ (справа):

Синя лінія - мінімальна ефективна користь (наприклад, чутливість ≥ 0.85) і нижче цього рівня алгоритм не приносить клінічної цінності.

Червона лінія — максимально допустимий ризик (наприклад, FNR ≤ 0.05), вище якого модель створює небезпеку для пацієнтів.

Зелена зона є ІВ, тобто діапазоном роботи алгоритму, де користь перевищує ризики.

Таким чином, ілюстрація підкреслює, що у фармакології баланс досягається між ефективністю та токсичністю препарату, а у медичних ШІ між корисністю алгоритму та ризиком помилок. Це візуальне пояснення, чому поняття «вікна» є універсальним

інструментом для оцінки користь–ризик лікарських засобів та ШІ.

Фактори, що впливають на ширину ІВ можуть включати пацієнтські особливості, різні юрисдикції, технічні аспекти, баланс даних, стійкість моделі до шуму.

Медичні установи часто стикаються з викликами, коли кілька алгоритмів застосовуються одночасно, наприклад, системи для діагностики, прогнозування та прийняття рішень [13]. Подібно до фармакологічних взаємодій препаратів, взаємодія кількох алгоритмів може змінювати ефективність і ризик кожного з них, що вимагає комплексної оцінки, аби зберегти роботу всієї системи в межах ІВ.

Таким чином, спільним для ТВ та ІВ є основний принцип «користь має переважати ризик» та має місце постійний моніторинг. Відмінним є характер користі й ризиків. Так для ліків це насамперед фармакологічні ефекти на організм, а для ШІ, якість даних, алгоритмічні помилки та інформаційна безпека. На практиці це означає, що лікарський засіб потребує ретельних клінічних випробувань, а ШІ, не лише медичної, а й технічної та етичної перевірки протягом усього життєвого циклу.

Сучасні методи оцінки користі та ризику клінічних алгоритмів ШІ і потреба їх доповнення концепцією «терапевтичного вікна»

Оцінка співвідношення користі та ризику є ключовим етапом упровадження будь-якої медичної технології, зокрема систем ШІ, що застосовуються для діагностики, прогнозування чи підтримки клінічних рішень. Наразі у сфері медичного ШІ використовуються різноманітні методи, від статистичних метрик ефективності до моделей клінічної корисності, аналізу прийняття рішень та оцінки потенційних наслідків помилок. Для розуміння цього контексту доцільно розглянути існуючі методології, які сьогодні застосовуються у сфері клінічного ШІ, їхні сильні сторони та обмеження, що формує підґрунтя для обґрунтування нової концепції «інформаційного вікна».

1. Регуляторні рамки (Guidance, Frameworks), які встановлені агентствами (FDA, EMA), описують стандарти оцінки ризиків й користі, з акцентом на надійність моделі як-от валідність, контекст використання, безпека, контроль за змінами (постмаркетинговий моніторинг). Такий підхід забезпечує юридичну і практичну основу, оскільки дозволяє стандартизувати вимоги, які допомагають виробникам, клінікам, регуляторам. На жаль, вони часто, не дають конкретного алгоритму порогів, які можуть бути надто широкими або абстрактними і приводять до проблем характеристики деяких клінічних випадків. FDA сама визнає, що жодна модель ШІ не є універсально точною. Її надійність має оцінюватися не «взагалі», а відповідно до конкретного контексту використання, тобто, у яких клінічних умовах, для якої мети, і з яким рівнем допустимого ризику модель застосовується [14].

2. Моделі Bayesian, які дозволяють працювати з невизначеністю. Вони не просто видають результат, а ще й оцінюють, наскільки впевнена модель у своїх прогнозах. Підходить навіть для ситуацій з обмеженою кількістю даних і є гнучкою з точки зору пріоритетів. Проте, такі моделі є складними і менш прозорими для користувачів (лікарів), тому результати можуть залежати від припущень [15].

3. Мультикритеріальний аналіз - це набір методів, які дозволяють ШІ приймати рішення, враховуючи кілька різних критеріїв одночасно, навіть якщо вони суперечать один-одному. Він використовує підхід, у якому користь та ризики представлені декількома критеріями, кожному з яких присвоюється вага (числовий коефіцієнт, що показує, наскільки важливий певний критерій у прийнятті рішення порівняно з іншими). Чим більша вага, тим більший вплив критерію на підсумковий вибір. Визначення такого показника є суб’єктивним процесом і потребує надійних даних усіх критеріїв, що може бути складним для клінічної практики [16].

4. Метод оцінки медичних технологій (ОМТ) застосовують для перевірки безпеки, ефективності, етичності та економічної доцільності систем ШІ у сфері охорони здоров’я. Часто спізнюється за

технологіями та не завжди адаптований для швидко змінюваних ШІ-моделей, оскільки має складність зі збором та аналізом даних [17].

5. Регуляторний підхід «Допустимий діапазон ризику» (Acceptable Risk Range, ARR), за яким технологія вважається прийнятною, якщо її ризики знаходяться в заздалегідь визначеному діапазоні [18]. Це «зона ризику», яка є прийнятною для організації або стандартів і не потребує негайного втручання.

Концепція ІВ пропонує аналітичне доповнення до цього підходу, оскільки не лише визначає діапазон прийнятності, а й кількісно описує співвідношення користі та ризику через показники ефективності (AUC, чутливість, специфічність) і безпеки (хибнопозитивні результати, індекс упередженості тощо). Таким чином, ІВ може розглядатися як математичне та методологічне розширення ARR, що дозволяє перейти від якісного (регуляторного) до кількісного (модельного) визначення безпечних меж функціонування ШІ у клінічних системах. Наша концепція ІВ сформована як розвиток ARR, що фокусується переважно на встановленні порогових значень безпеки.

Таким чином, ці дані свідчать про те, що поточні методології вже мають такі важливі складові як продуктивність (чутливість, точність тощо), безпеку, контекст використання, вплив на систему та етику. Водночас спостерігається і дисбаланс, оскільки користь часто оцінюється більш кількісно та формально, тоді як ризики іноді лише описові або оцінюються поверхово. Недостатня є й стандартизація порогів, за якими значення метрик користі чи ризику є прийнятними, що утворює невизначеність. Такі аспекти як кібербезпека та адверсаріальний ризик також часто не інтегруються повністю у відповідний аналіз, тому після випуску моделі існує потреба в життєвому циклі, який складає моніторинг, оновлення та перевірка на реальних даних.

Теоретико-математична модель ІВ в оцінці користі та ризику систем ШІ

Визначення та формалізація проблеми. Нехай алгоритм медичного ШІ при заданому робочому порозі x має дві основні функції характеристик:

B(x) - функція користі (benefit), що відображає клінічну ефективність (наприклад, чутливість, AUC або композитна метрика).

R(x) - функція ризику (risk), що відображає небажані наслідки (наприклад, частота хибнопозитивних/хибнонегативних результатів, індекси упередженості, або комбінований ризик). ІВ (IW) визначається як множина робочих точок (порогів) x із виконанням наступних умов:

$$IW = \{x \in X | B(x) \geq MEB \wedge R(x) \leq MRB\}$$

де: MEB (Minimal Effective Benefit) - мінімальний допустимий рівень користі, визначений клінічно (нижня межа); MRB (Maximum Reasonable Burden) - максимальний допустимий рівень ризику, визначений клінічно/регуляторно (верхня межа). Для чисельного моделювання часто використовують вектор порогів x_i та дискретні оцінки B(x_i) та R(x_i).

Розрахунок MEB та MRB

$$MEB = min\{B(x) | B(x)) \geq B_{crit}\}$$

де: B_{crit} – клінічно значущий поріг (наприклад, чутливість ≥ 0.8).

$$MRB = \text{man}\{R(x) | R(x)) \geq R_{crit}\}$$

де R_{crit} – гранично допустимий ризик (наприклад, хибнонегативи ≤ 5%).

Відношення користі до ризику (Benefit-Risk Ratio)

$$BRR(x) = \frac{B(x)}{R(x)}$$

і вважати, що алгоритм знаходиться у безпечному інформаційному вікні, якщо:

$$\text{BRR(x)} \geq \theta$$

θ порогове значення (наприклад, ≥ 2, тобто користь удвічі перевищує ризик).

Ширина інформаційного вікна (аналог «ширини терапевтичного вікна)**:**

$$Width(IW) = MRB - MEB.$$

Якщо має місце вузьке вікно, алгоритм вимагає постійного моніторингу, якщо широке вікно, то алгоритм є більш стійким і безпечним.

Практична реалізація досягається наступним чином: обчислюються криві *B(x)* та *R(x)* залежно від порогу алгоритму (ROC-аналіз); визначаються значення *MEB* та *MRB* за заздалегідь встановленими клінічними критеріями; будуються діаграми (аналог фармакокінетичної кривої), що дозволяє візуалізувати зону прийнятності. Отже, метод поєднує аналітичні формули (*B/R*) з графічною візуалізацією (вікно), що робить його зручним і для клініцистів, і для регуляторів.

Математичний приклад використання концепції ІВ до відомої медичної моделі (алгоритму для виявлення діабетичної ретинопатії) [19]. У цій статті були наведені операційні точки алгоритму на валідаційних наборах: sensitivity ≈ 0.975 та specificity ≈ 0.934 (одна з валідаційних множин, у статті також наведено інші точки і ROC-криву). Ці значення ми використаємо як приклад конкретного робочого порогу алгоритму.

Визначення функцій користі й ризику.

Візьмемо просту, інтуїтивну формалізацію: *B(x)* - користь у точці x (обернено: краща - вища). Для прикладу приймемо *B(x)* = sensitivity при порогі *x*. *R(x)* - ризик у точці *x*. Для ілюстрації візьмемо ризик пропуску хвороби (critical у скринінгу) як false negative rate:

$$R(x) = FNR(x) = 1 - sensitivity(x)$$

За іншої клінічної задачі можна вибрати *R(x)* = *FPR* або комбінований ризик; вибір залежить від клінічних пріоритетів.

Приклад клінічно-обґрунтованих порогів. Якщо експерти домовилися про такі критерії для скринінгової моделі як *MEB*: B(x)≥0.90 (тобто мінімальна чутливість 90 %), а *MRB R(x)*≤0.05 (тобто *FNR* ≤ 5 %, тобто чутливість ≥ 95 %). Ці числа - ілюстративні, але реалістичні для задачі скринінгу діабетичної ретинопатії, де пропуски (*FN*) несуть серйозні наслідки.

Перевірка конкретного робочого порогу (числовий приклад). Публічно наведений робочий пункт алгоритму: sensitivity=0.975. Отже: Обчислюємо *R* (*FNR*):

$$R = 1 - sensitivity = 1 - 0.975 = 0.025$$

Порівнюємо з порогами:

- чи *B*≥*MEB*? 0.975≥0.90 – так;
- чи *R* ≤*MRB*? 0.025≤0.05 – так.

Отже, при цьому робочому порозі алгоритм знаходиться всередині ІВ за нашими умовами.

Обчислимо BRR як відповідне співвідношення:

$$\mathrm{BRR(x)} = \frac{\mathrm{B(x)}}{\mathrm{R(x)}} = \frac{0.975}{0.025} = 39$$

У цьому випадку користь (чутливість) у 39 разів перевищує ризик (FNR) за цією простою метрикою і є дуже «вигідне» співвідношення для задачі скринінгу. Зверніть увагу: *BRR* тут допоміжна числова інтерпретація; його поріг θ треба визначити клінічно. Наприклад, θ≥2 означатиме «користь удвічі перевищує ризик».

Повний розрахунок інформаційного вікна при наявності ROC/threshold data.

Щоб обчислити повне ІВ (тобто інтервал порогів *x,* для якого одночасно виконуються вимоги), потрібна крива залежності sensitivity(*x*), specificity(*x*) або інші метрики від порогу.

Процедура:

1. Отримати дискретну множину пар:

$$\{x_i, sensitivity(x_i), FNR(x_i), FPR(x_i)\}$$

або безперервну ROC функцію.

2. Задати *MEB* і *MRB* (клінічно обґрунтовано).

3. Обчислити множину порогів:

$$\mathrm{IW} = \left\{ \begin{matrix} \mathrm{x_i} | \mathrm{sensitivity(x_i)} \geq \\ \mathrm{MEB} \wedge \mathrm{FNR(x_i)} \leq \mathrm{MRB} \end{matrix} \right\}$$

4. Нижня межа вікна - мінімальний x у цьому наборі; верхня межа - максимальний x. Якщо x є межею від 0 до 1, то ширина вікна:

$$Wigth(IW) = x_{max} - x_{min}.$$

Так, якщо зі знімків ROC ми знаходимо, що для порогів x > [0.12,0.48] виконані обидві умови, то ІВ = [0.12, 0.48], ширина = 0.36.

Наведений числовий приклад показує лише одну операційну точку алгоритму (чутливість ≈ 0.975) і тому демонструє, що ця точка лежить у ІВ за заданими порогами. Щоб знайти повний інтервал, необхідні значення *sensitivity(x)* та *FNR(x)* для множини порогів (тобто повна ROC/межа у таблиці), яку автори наводять у статті або яку можна отримати з валідаційного набору.

У задачах, де важливі й інші аспекти (наприклад, надмірні хибнопозитиви, упередження стосовно підгруп пацієнтів, економічні витрати), *R(x)* краще формалізувати як композиційний ризик:

$$R(x) = \omega_1 FNR(x) + \omega_2 FNR(x) + \omega_3 BiasIndex(x) \ldots$$

з вагами ωі, визначеними клінічно/регуляторно. Тоді умова R(x)≤MRB стає багатовимірною і відображає компроміс між різними типами несприятливих наслідків.

Порівняльна оцінка параметрів користі та ризику ІВ і традиційними методами аналізу медичних алгоритмів ШІ

Нами проведено аналіз оцінки ефективності та безпеки використання ІВ, як нового методу, порівняно з традиційними ШІ-алгоритмами з точки зору точності, надійності та практичної застосовності в клінічних умовах. Таким чином забезпечується науково обґрунтована основа прийняття рішень щодо впровадження ШІ у медичну практику. Оскільки оцінка співвідношення «користь–ризик» у сфері медичних ШІ має багатовимірний характер і охоплює не лише технічні характеристики алгоритмів, але й клінічні, етичні та соціальні чинники, їх для наочності аналізу можна згрупувати у три ключові сфери, які є основними щодо їх впровадження в практику (рис. 2).

Ліве коло визначає регуляторні рамки, які визначають продуктивність моделі (точність, ROC, чутливість та специфічність). Тут головне завдання розробників довести, що алгоритм дійсно виконує свою клінічну задачу у визначених умовах. Для регулятора важливо бачити, що система працює на певному рівні безпеки та ефективності перед впровадженням. Сильними сторонами є чіткі кількісні метрики продуктивності та належна оцінка конкретного клінічного контексту використання. Втім, недостатня увага приділена до кібербезпеки, не враховані системні чи соціальні наслідки та не покриті етичні ризики.

Праве коло характеризує безпеку та етику для пацієнтів, включно з помилковими рішеннями та упередженістю алгоритмів. Особлива увага приділяється адверсаріальним ризикам, коли дані чи алгоритми можуть бути навмисно або випадково спотворені. Враховані етичні аспекти справедливості та прозорості, але недостатньо охоплені показники продуктивності алгоритму та економічних або системних наслідків.

Нижнє коло відповідає за методи аналізу та моделювання. Воно забезпечене баєсівськими моделями, MCDA, HTA тощо, які працюють з невизначеністю, економічними та системними наслідками, масштабованістю та якістю життя. У той же час вони не враховують безпеку пацієнта у клінічному сенсі та обмежена увага до кіберризиків і етики.

Зони перетину складаються із трьох частин взаємодії:

1. Регуляторні та аналіз, які частково поєднують продуктивність із економічними аспектами та контекстом використання.

2. Регуляторні та безпека відповідають за клінічну безпеку із вимогами до продуктивності.

3. Аналіз та безпека дозволяють враховувати ризики разом із невизначеністю та системним впливом.

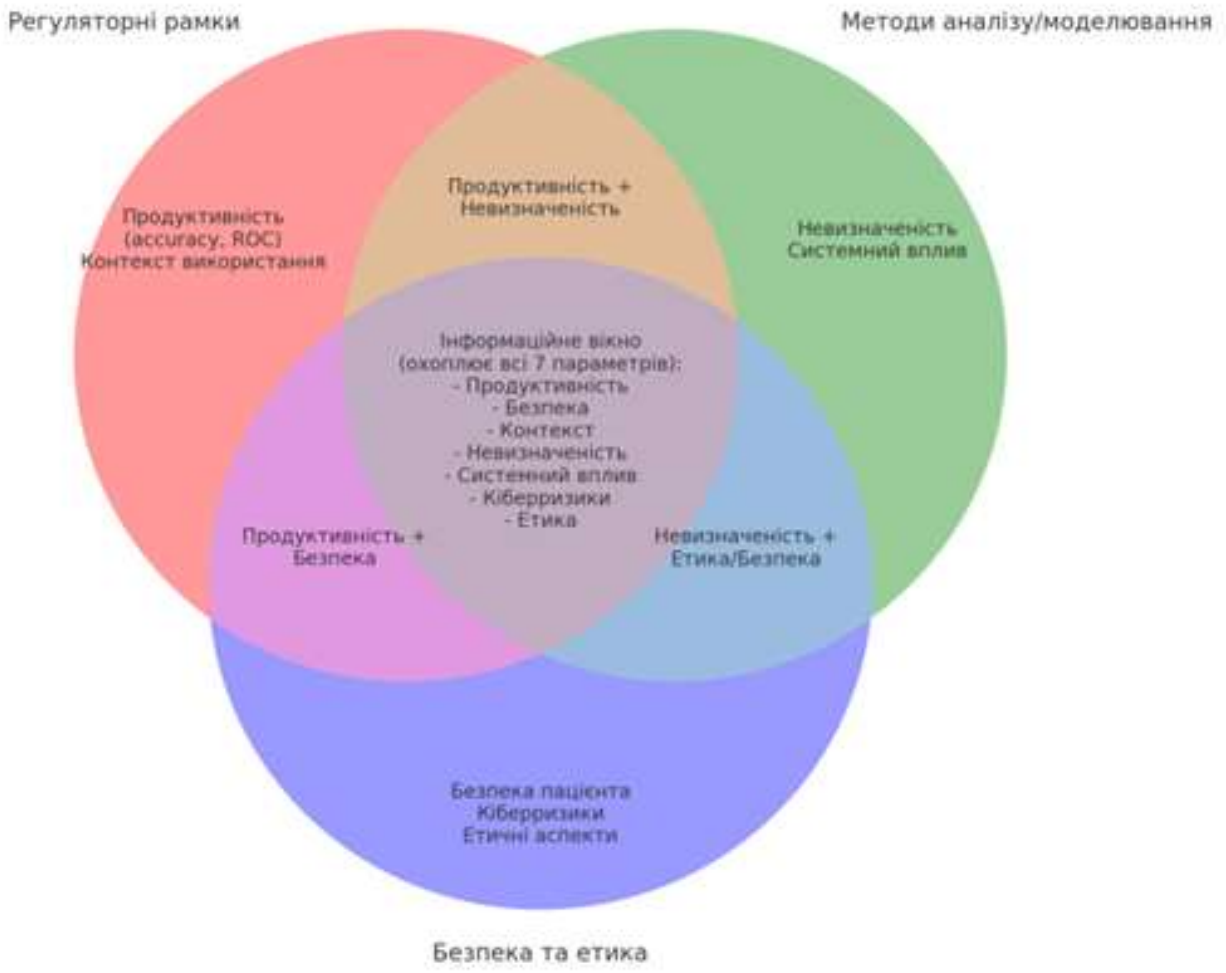

Рис. 2. Параметри «користь-ризик» у медичних ШІ, та інтеграція їх в концепції ІВ

Отож, сучасні підходи є фрагментованими, де кожен охоплює лише частину проблеми. За таких обставин, ІВ постає як цілісна інтеграційна концепція, яка дозволяє бачити повну картину й приймати рішення на основі комплексної оцінки користь–ризик. Тому ІВ представлено центральною зоною, де зливаються всі ключові аспекти, продуктивність моделі, безпека пацієнта, клінічний контекст використання, невизначеність, системний і економічний вплив, кіберризики та етичні аспекти, що власне і складає ІВ. Його центральність не є поки що емпірично доведеною в повному сенсі, однак прогнозно обґрунтована через системну аналогію з терапевтичним вікном, математичне моделювання зон оптимальності та відповідність вимогам сучасних регуляторних рамок (FDA SaMD, EU AI Act). Вона також підтверджена наступними доповненнями:

1. Теоретико-аналогічна основа. Модель побудована за аналогією до «терапевтичного вікна» у фармакології, де існує оптимальний діапазон дії препарату між неефективною і токсичною дозою. У ШІ це переноситься на інформаційне навантаження системи між недоінформованістю (високий ризик помилки) та перенасиченням даними (зростанням когнітивного шуму та ризиків).

2. Емпірична спостережуваність. Аналіз практичних кейсів (наприклад, MIMIC-III, Med-PaLM, BioGPT) показує, що системи ШІ мають діапазон стабільної продуктивності, за межами якого зростає кількість помилкових рішень або упереджених результатів.

3. Прогностичне моделювання. Математичні симуляції (на основі регресійних і диференційних моделей) дозволяють оцінити, при яких параметрах (розмір тренувальних даних, рівень шуму,

систематична похибка, довірчий інтервал) система переходить із зони оптимальної роботи у зону ризику.

Подібно до лікарських засобів, системи ШІ медичного призначення характеризуються власним балансом користі та ризику, який визначається сукупністю технічних, клінічних і контекстуальних факторів. Саме тому практичне кількісне визначення параметрів ІВ та порівняння їх зі стандартними методами різних моделей медичних алгоритмів набуває особливого значення, оскільки дозволяє ідентифікувати межі безпечного та ефективного використання ШІ у різних клінічних контекстах, виявити критичні чинники, що створюють передумови для стандартизації оцінки користі й ризику між різними класами. Задля створення такого порівняння (таблиця 1) нами використано відкриті дані [20-24], які дозволяють ретроспективно реконструювати цю аналогію.

Таблиця 1. Порівняння оцінки користі–ризику медичних ШІ методами стандартної метрики та ІВ

Модель / джерело	Метод оцінки	Основні показники	Висновок за стандартними метриками	Висновок за методом «Інформаційного вікна»	Ключова відмінність
Gulshan et al.,[20]	ROC/AUC, чутливість, специфічність	AUC=0.99, Sens=0.975, Spec=0.934	Висока ефективність, мінімальний ризик	MEB ≥0.9, MRB ≤0.05 → ІВ широке	Обидві оцінки збігаються, але ІВ дає діапазон прийнятності, а не просто оцінку.
Esteva et al., [21]	AUC, точність, повнота	AUC=0.96, Sens≈0.91	Модель ≈ рівень дерматології	MEB=0.85, MRB=0.1 → ІВ середнє	ІВ показує, що навіть при високій точності ризик FP залишається клінічно значущим.
Rajpurkar et al., [22]	AUC, показник F1	AUC=0.76, Sens=0.90, Spec=0.58	Модель помірна	MEB=0.85, MRB=0.15 → ІВ вузьке	ІВ сигналізує, що модель нестабільна для клінічної інтеграції.
Ribeiro et al., [23]	AUC для кожної патології	0.76–0.84	Залежить від діагнозу	Для кількох патологій MEB < MRB → ІВ відсутнє	ІВ показує, де саме модель виходить за межі «допустимого ризику».
DeepMind, [24]	AUC, коефіцієнт Каппа	AUC=0.99	Найкращий рівень	MEB=0.95, MRB=0.05 → ІВ вузьке, але стабільне	ІВ виявляє критичну зону, де FP/ FN потребують клінічного моніторингу.

Примітка: ROC - крива робочих характеристик приймача; AUC - площа під ROC-кривою.

Порівняльний аналіз продемонстрував, що метод оцінки користі та ризику на основі концепції ІВ є не просто альтернативою, а суттєвим розширенням

традиційних статистичних метрик, таких як AUC, чутливість та специфічність. Стандартні показники дозволяють лише кількісно оцінити точність алгоритму, проте не відображають клінічної прийнятності моделі в умовах реальної варіабельності даних та ризику хибних результатів. Концепція IB, натомість, вводить додатковий вимір, як-от діапазон допустимої роботи алгоритму між мінімальною ефективністю (MEB) та максимально допустимим ризиком (MRB), який визначається не лише статистично, а й клінічно чи регуляторно.

У наведених прикладах видно, що для високоточних систем (Google AI чи DeepMind) межі IB залишаються широкими або стабільними, що свідчить про надійну інтеграцію між користю й ризиком. Водночас для моделей із нижчою специфічністю (CheXNet, ChestX-ray14) IB різко звужується або взагалі зникає, що сигналізує про потребу у додатковій валідації чи адаптації до конкретного клінічного контексту. Отже, IB дозволяє перейти від абстрактної оцінки «точності» до реальної оцінки клінічної безпеки, встановлюючи межі, за якими алгоритм виходить із зони прийнятного ризику. Такий підхід створює методологічний зв'язок між науковим аналізом ефективності та регуляторним принципом ARR, який використовується FDA та EMA для визначення допустимих меж ризику у клінічних технологіях. Відтак, IB не замінює традиційних метрик, а доповнює їх системною рамкою, що дозволяє одночасно оцінювати аналітичну точність, клінічну користь і безпекові ризики, у кількісно визначеному форматі, придатному для стандартизації та регуляторного використання.

Висновки

1. Дослідження має концептуально-теоретичний характер і формує основу для подальших прикладних робіт, зокрема, побудови емпіричних моделей розрахунку IB у різних типах медичних алгоритмів (діагностичних, прогностичних, підтримки прийняття рішень).

2. Запропоновано нову методологічну концепцію «інформаційне вікно», яка є функціональним аналогом терапевтичного вікна у фармакології та покликана забезпечити системне оцінювання співвідношення користі та ризику у медичних системах ШІ. На відміну від традиційних метрик (AUC, чутливість, специфічність), IB дозволяє одночасно інтегрувати показники ефективності та безпеки в єдину кількісну модель.

3. Порівняльний аналіз показав методологічну спорідненість між терапевтичним вікном лікарських засобів і межами прийнятного ризику у ШІ, зокрема через спільну логіку оптимального балансу між максимальною клінічною користю (MEB) та мінімальним ризиком (MRB). Ці параметри у ШІ можуть виступати аналітичними аналогами дозування та токсичності у фармакології.

4. Концепція IB узгоджується з міжнародними регуляторними принципами системи «ARR» (FDA, EMA), Європейський акт про ШІ (EU AI Act), рекомендації OECD та етичні стандарти IEEE. Вона може слугувати науковим підґрунтям для створення єдиних критеріїв прийнятності моделей ШІ у клінічній практиці.

5. Розроблена теоретико-математична модель IB формалізує взаємозв'язок між чутливістю, специфічністю, часткою хибнопозитивних/хибнонегативних результатів, а також етичними та контекстними факторами використання моделі (Credibility/Context of Use). Це створює можливість побудови універсальної шкали безпечності алгоритмів у різних медичних задачах.

6. Порівняльна оцінка моделей показала, що IB краще відображає клінічну прийнятність, ніж традиційні метрики: у випадках високої точності, але значного ризику помилкових класифікацій (*FP*, *FN*), *IB* дозволяє виявити зону критичних меж, що є невидимою для стандартних оцінок продуктивності.

7. Практична значущість концепції IB полягає в її здатності слугувати мостом між наукою, клінічною практикою та регуля-

торною політикою, що допомагає стандартизації систем ШІ.

Література

1. Ажажа, М., Венгер, О., & Фурсін, О. (2023). Концепція цифрового маркетингу 4.0: еволюція, характеристика, типологія. Humanities Studies: Collection of Scientific Papers / за ред. В. Воронкової. Запоріжжя: Publishing house “Helvetica”, 14(91), 135–147. https://doi.org/10.32782/hst-2023-14-91-16

2. Amann, J., Blasimme, A., Vayena, E., et al. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20, 310. https://doi.org/10.1186/s12911-020-01332-6

3. Головенко Н. Я (2004) Физико-химическая фармакология. Одесса. 2004 Феникс. 740 с.

4. Muller, P. Y., & Milton, M. N. (2012). The determination and interpretation of the therapeutic index in drug development. Nature Reviews Drug Discovery, 11(10), 751–761. https://doi.org/10.1038/nrd3801

5. World Health Organization. (2025, March 25). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models (98 p.). https://www.who.int/publications/i/item/9789240084759

6. Головенко, М. Я. (2012). Фармакокінетичний моніторинг — запорука раціональної фармакотерапії.
Журнал НАМН України, 18(4), 440–445.

7. Chen, F., et al. (2024). Unmasking bias in artificial intelligence: A systematic review of bias detection and mitigation strategies in electronic health record-based models. Journal of the American Medical Informatics Association, 31(5), 1172–1183. https://doi.org/10.1093/jamia/ocae060

8. Шевченко, А. І. (ред.) (2023). Стратегія розвитку штучного інтелекту в Україні. Київ: Видавництво «Торпеда». 306 с.

9. Висоцький, А. А., Суріков, О. О., & Василюк-Зайцева, С. В. (2023). Розвиток штучного інтелекту в сучасній медицині. Український медичний часопис, 2(154), 1–4.
https://doi.org/10.32471/umj.1680-3051.154.2412213

10. Свінціцький, А. В., Климова, В. В., & Сендецький, С. С. (2024). Використання штучного інтелекту в медицині, хірургії, стоматології, онкології. Клінічна онкологія, 14(3[55]), 1–4. https://doi.org/10.32471/clinicaloncology.2663-466X.56-4.33692

11. Wiens, J., Saria, S., Sendak, M., et al. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

12. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

13. Soenksen, L., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., Wiberg, H., Li, M. L., Fuentes, I., & Bertsimas, D. (2022). Integrated multimodal artificial intelligence framework for healthcare applications. npj Digital Medicine, 5, 149. https://doi.org/10.1038/s41746-022-00689-4

14. U.S. Food and Drug Administration. (2024). FDA proposes framework to advance credibility of AI models used for drug and biological product submissions. https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions

15. Bai, T., Lan, H., & Tiwari, R. (2021). Bayesian approaches to benefit–risk assessment for diagnostic tests. Journal of Biopharmaceutical Statistics, 31(4), 541–558.
https://doi.org/10.1080/10543406.2021.1931272

16. Vuong, Q., Metcalfe, R. K., Harari, O., Mills, E. J., & Park, J. J. H. (2025, February 12). Benefit–risk assessment of medical products using Bayesian multi-criteria augmented decision analysis for clinical development. medRxiv.
https://doi.org/10.1101/2023.08.31.23294918

17. Farah L, Borget I, Martelli N, Vallee A. Suitability of the Current Health Technology Assessment of Innovative Artificial Intelligence-Based Medical Devices: Scoping Literature Review. J Med Internet Res 2024; 26: e51514. doi: 10.2196/51514.

18. Fraser, H., & Bello Villarino, J.-M. (2024). Acceptable risks in Europe’s proposed AI Act: Reasonableness and other principles for deciding how much risk management is enough. European Journal of Risk Regulation, 15(2), 431–446.
https://doi.org/10.1017/err.2023.57

19. Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

20. Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

21. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, 115–118. https://doi.org/10.1038/nature21056

22. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). Radiologist-level pneumonia detection on chest X-rays with deep learning: CheXNet. arXiv preprint arXiv:1711.05225.
https://doi.org/10.48550/arXiv.1711.05225

23. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. arXiv preprint arXiv:1602.04938. https://arxiv.org/abs/1602.04938

24. Yim, J., Chopra, R., De Fauw, J., & Ledsam, J. (2020). Using AI to predict retinal disease progression. Nature. DeepMind.
https://www.nature.com/articles/s41586-020-2974-4

References

1. Azhazha, M., Venher, O., & Fursin, O. (2023). Kontseptsiia tsyfrovoho marketynhu 4.0: evoliutsiia, kharakterystyka, typolohiia. Humanities Studies: Collection of Scientific Papers / za red. V. Voronkovoi. Zaporizhzhia: Publishing house “Helvetica”, 14(91), 135–147. https://doi.org/10.32782/hst-2023-14-91-16

2. Amann, J., Blasimme, A., Vayena, E., et al. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20, 310. https://doi.org/10.1186/s12911-020-01332-6

3. Holovenko N.Ya (2004) Fyzyko-khymycheskaia farmakolohyia. Odessa. 2004 Fenyks. 740 s.

4. Muller, P. Y., & Milton, M. N. (2012). The determination and interpretation of the therapeutic index in drug development. Nature Reviews Drug Discovery, 11(10), 751–761. https://doi.org/10.1038/nrd3801

5. World Health Organization. (2025, March 25). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models (98 p.). https://www.who.int/publications/i/item/9789240084759

6. Holovenko, M. Ya. (2012). Farmakokinetychnyi monitorynh — zaporuka ratsionalnoi farmakoterapii. Zhurnal NAMN Ukrainy, 18(4), 440–445.

7. Chen, F., et al. (2024). Unmasking bias in artificial intelligence: A systematic review of bias detection and mitigation strategies in electronic health record-based models. Journal of the American Medical Informatics Association, 31(5), 1172–1183. https://doi.org/10.1093/jamia/ocae060

8. Shevchenko, A. I. (red.) (2023). Stratehiia rozvytku shtuchnoho intelektu v Ukraini. Kyiv: Vydavnytstvo «Torpeda». 306 s.

9. Vysotskyi, A. A., Surikov, O. O., & Vasyliuk-Zaitseva, S. V. (2023). Rozvytok shtuchnoho intelektu v suchasnii medytsyni. Ukrainskyi medychnyi chasopys, 2(154), 1–4. https://doi.org/10.32471/umj.1680-3051.154.2412213

10. Svintsitskyi, A. V., Klymova, V. V., & Sendetskyi, S. S. (2024). Vykorystannia shtuchnoho intelektu v medytsyni, khirurhii, stomatolohii, onkolohii. Klinichna onkolohiia, 14(3[55]), 1–4. https://doi.org/10.32471/clinicaloncology.2663-466X.56-4.33692

11. Wiens, J., Saria, S., Sendak, M., et al. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

12. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

13. Soenksen, L., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., Wiberg, H., Li, M. L., Fuentes, I., & Bertsimas, D. (2022). Integrated multimodal artificial intelligence framework for healthcare applications. npj Digital Medicine, 5, 149. https://doi.org/10.1038/s41746-022-00689-4

14. U.S. Food and Drug Administration. (2024). FDA proposes framework to advance credibility of AI models used for drug and biological product submissions. https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions

15. Bai, T., Lan, H., & Tiwari, R. (2021). Bayesian approaches to benefit–risk assessment for diagnostic tests. Journal of Biopharmaceutical Statistics, 31(4), 541–558. https://doi.org/10.1080/10543406.2021.1931272

16. Vuong, Q., Metcalfe, R. K., Harari, O., Mills, E. J., & Park, J. J. H. (2025, February 12). Benefit–risk assessment of medical products using Bayesian multi-criteria augmented decision analysis for clinical development. medRxiv. https://doi.org/10.1101/2023.08.31.23294918

17. Farah L, Borget I, Martelli N, Vallee A. Suitability of the Current Health Technology Assessment of Innovative Artificial Intelligence-Based Medical Devices: Scoping Literature Review. J Med Internet Res 2024;26:e51514. doi: 10.2196/51514.

18. Fraser, H., & Bello Villarino, J.-M. (2024). Acceptable risks in Europe’s proposed AI Act: Reasonableness and other principles for deciding how much risk management is enough. European Journal of Risk Regulation, 15(2), 431–446. https://doi.org/10.1017/err.2023.57

19. Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

20. Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

21. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, 115–118. https://doi.org/10.1038/nature21056

22. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). Radiologist-level pneumonia detection on chest X-rays with deep learning: CheXNet. arXiv preprint arXiv:1711.052

23. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. arXiv preprint arXiv:1602.04938. https://arxiv.org/abs/1602.04938

24. Yim, J., Chopra, R., De Fauw, J., & Ledsam, J. (2020). Using AI to predict retinal disease progression. Nature. DeepMind. https://www.nature.com/articles/s41586-020-2974-4

The article has been sent to the editors 15.12.25.
After processing 20.12.25.
Submitted for printing 30.12.25.

Copyright under license CCBY-SA4.0.